

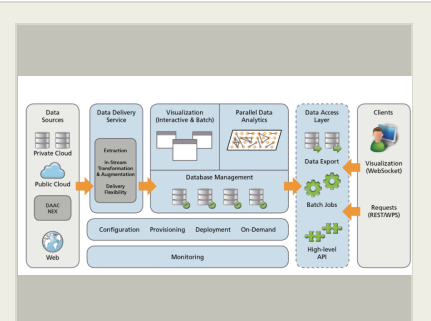
Prototyping agile production, analytics and visualization pipelines for big-data on the NASA Earth Exchange (NEX)

Completed Technology Project (2015 - 2017)



Project Introduction

The goal of this project is to develop capabilities for an integrated petabyte-scale Earth science product development, production and collaborative analysis environment. We will deploy this environment within the NASA Earth Exchange (NEX) and OpenNEX in order to enhance existing science data production pipelines in both high-performance computing (HPC) and cloud environments. Bridging of HPC and cloud is a fairly new concept under active research. This system will significantly enhance the ability of the scientific community to accelerate transformation of Earth science observational data from NASA missions, model outputs and other sources into science data products and facilitate collaborative analysis of the results. We propose to develop a web-based system that seamlessly interfaces with both high-performance computing (HPC) and cloud environments, providing tools that enable science teams to develop and deploy large-scale data processing pipelines, perform data visualization, provenance tracking, analysis and QA of both the production process and the data products, and enable sharing results with the community. In terms of the NRA, the project is proposed under 'Data-Centric Technologies' category. The HPC component will interface with the NASA Earth Exchange (NEX), a collaboration platform for the Earth science community that provides a mechanism for scientific collaboration, knowledge and data sharing together with direct access to over 1PB of Earth science data and 10,000-cores processing system. The cloud component will interface with NASA OpenNEX ' a cloud-based component of NEX. The project aligns well with number of goals of 'NASA's Plan for a Climate-Centric Architecture' and will be capable of supporting number of missions such as LDCM, OCO-2, or SMAP. There will be immediate benefit to number of existing and upcoming projects. The WELD (Web Enabled Landsat Data) project sponsored by NASA MEASUREs program will benefit immediately through improved production and QA monitoring capabilities as well as more efficient execution. There are also a number of projects that are ready to build on the WELD results. First of them is NASA GIBS, a core EOSDIS component, that requires to deliver native resolution imagery from WELD (this will be about 5PB production system). There are also science projects that hope to build on WELD results by implementing MODIS algorithms such as FPAR/LAI using the high-resolution Landsat data. In order to demonstrate the capabilities of the system, we will deploy a prototype on the existing NEX Landsat WELD processing system ' a complex 30-stage pipeline, which delivers derived vegetation products by processing over 1.5PB of data. The project will be developed in several stages each addressing separate challenge ' workflow integration, parallel execution in either cloud or HPC environments and big-data analytics and visualization. We will first develop the capability and best practices to assist science teams with integration of their large-scale processing pipelines with the workflow system. We will continue with enabling users to launch seamless data production on either cloud or HPC environments, while tracking the data and process provenance. This effort will be based on previous ESTO-funded activities. Finally, we will integrate the system with web-base visualization



NEX Big Data Pipeline

Table of Contents

Project Introduction	1
Organizational Responsibility	1
Anticipated Benefits	2
Primary U.S. Work Locations and Key Partners	2
Project Management	2
Technology Maturity (TRL)	2
Technology Areas	2
Images	3
Target Destination	3

Organizational Responsibility

Responsible Mission Directorate:

Science Mission Directorate (SMD)

Lead Center / Facility:

NASA Headquarters (HQ)

Responsible Program:

Advanced Information Systems Technology

Prototyping agile production, analytics and visualization pipelines for big-data on the NASA Earth Exchange (NEX)

Completed Technology Project (2015 - 2017)

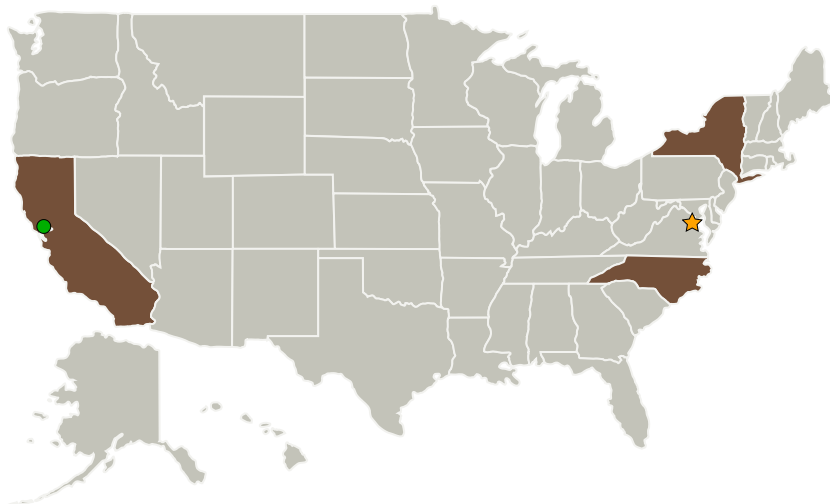


tools to enable efficient big-data visualization and analytics of the results. The period of performance of the project is two years and we have estimated the possible beginning for March 2015. However, the exact start date is not critical for this project and it can be readily adjusted. We have estimated the entry TRL of the efforts at 4 and we will deliver a system with exit TRL of 6. The detailed TRL justification is provided in the proposal.

Anticipated Benefits

SMAP

Primary U.S. Work Locations and Key Partners



Organizations Performing Work	Role	Type	Location
★ NASA Headquarters(HQ)	Lead Organization	NASA Center	Washington, District of Columbia
● Ames Research Center(ARC)	Supporting Organization	NASA Center	Moffett Field, California
Kitware, Inc.	Supporting Organization	Industry	Clifton Park, New York

Project Management

Program Director:

Pamela S Millar

Program Manager:

Jacqueline J Le Moigne

Principal Investigator:

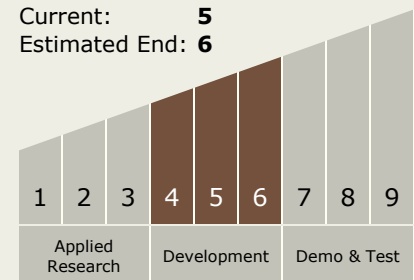
Aashish Chaudhary

Co-Investigators:

Petr Votava
Andrew R Michaelis
Charles P Reynolds
Ramakrishna R Nemani
Katherine Osterdahl
Jeffrey Baumes

Technology Maturity (TRL)

Start: 4
Current: 5
Estimated End: 6



Technology Areas

Primary:

- TX11 Software, Modeling, Simulation, and Information Processing
 - TX11.4 Information Processing

Continued on following page.

Prototyping agile production, analytics and visualization pipelines for big-data on the NASA Earth Exchange (NEX)

Completed Technology Project (2015 - 2017)



Primary U.S. Work Locations

California

New York

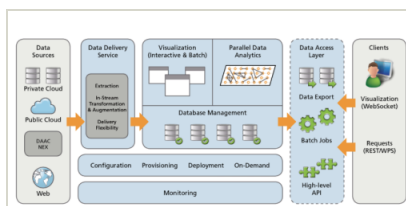
North Carolina

Technology Areas (cont.)

↳ TX11.4.4 Collaborative Science and Engineering

Target Destination Earth

Images



Untitled

NEX Big Data Pipeline

(<https://techport.nasa.gov/image/16563>)